DOCUMENT RESUME

ED 235 195 TM 830 590

AUTHOR Bangert, Robert L.; And Others
TITLE Effects of Coaching Programs on Aptitude Test Scores.
PUB DATE Apr 83
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Aptitude Tests; *Meta Analysis; Pretesting; *Program Effectiveness; Research Needs; Research Problems; *Scores; *Test Coaching
IDENTIFIERS *Effect Size; *Scholastic Aptitude Test

ABSTRACT

A quantitative method, meta-analysis, is used to integrate research findings from a representative group of 38 coaching studies. Because different studies reported results on different scales, effect size was used to transform all results to a common metric. This meta-analysis showed that there are two distinct literatures on the effectiveness of coaching programs. The first is on the Scholastic Aptitude Test (SAT), and reports small effects from coaching. The second covers other aptitude tests and shows that coaching programs can have substantial effects. Studies that used a pretest yielded larger estimates of pure coaching effects than did other studies. This indicates that a pretest may be an important component in any program designed to prepare students for aptitude tests. Results support the conclusion that variation in study findings is only modestly predictable from study characteristics. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED235195

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Effects of Coaching Programs on Aptitude Test Scores

Robert L. Bangert, James A. Kulik, & Chen-Lin C. Kulik

Center for Research on Learning and Teaching

The University of Michigan

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. Bangert

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A Paper Presented at the Annual Meeting

Of the American Educational Research Association

Montreal, April 1983

TM 830 590

Aptitude testing is a multimillion dollar industry that plays an important role in American education. Every year millions of students take such tests as the Graduate Record Examination, the Scholastic Aptitude Test, and the Law School Admissions Test, and their lives are affected by the results. In recent years a "coaching" industry has grown up in the shadow of the testing establishment. This satellite industry offers "coaching" and "crash" courses to help students improve their chances of scoring high on admissions and aptitude tests. The coaching industry is made up of at least 150 independent firms, and it offers services for 50,000 students annually.

The testing and coaching industries embody different beliefs about aptitude testing. According to the testers, aptitude tests measure capacities that are developed gradually from in-school and out-of-school experiences--capacities that are not likely to be changed significantly by short-term coaching (College Entrance Examination Board, 1968, p. 8). The coaching industry, on the other hand, maintains that aptitude test scores can be raised by such practices as test-familiarization, drill and practice, instruction in test-taking strategy, and highly focussed content teaching.

It is difficult to decide on the basis of individual research studies which of these views is more reasonable. Studies of coaching have been carried out in different settings, with different experimental designs, and with different results. Some studies have produced results that support the testing industry's view on the modifiability of aptitude test scores, whereas other studies produced results that support the coaching industry's view.

Reviews of coaching studies have not resolved the controversy. The first reviews were written in England and supported the conclusion that coaching has a significant influence on test performance. In one of the best of the British reviews, Vernon (1954) reported that the average effect of coaching and practice was to increase IQ scores by 8 to 9 points, or by about .6 standard deviations. Vernon pointed out that such an effect could be achieved in a remarkably short time, usually between 3 and 9 hours. More recent reviews of coaching studies have focussed on the widely used SAT. These reviews have generally emphasized the futility of coaching. The trustees of the College Board, for example, stated that the average increase to be expected from an intensive coaching program of perhaps 15 to 20 hours would be .10 standard deviations (College Entrance Examination Board, 1968).

There are at least two reasons for the inconsistency in conclusions about the effects of coaching: (1) reviewers have not examined the same studies; and (2) reviewers have

not analyzed the accumulated study results with quantitative and statistical methods. Our study was meant to overcome these limitations. It used a quantitative method--meta-analysis--to integrate research findings from a large and representative group of coaching studies.

Method

The first step in our meta-analysis was to collect the studies. We located in all 35 separate reports on 38 different studies. These reports came from journal articles, dissertations, and ERIC documents.

The coaching procedures and tests in the 38 studies were of several different types. We first classified the studies according to these program and test features (Table 1). The first of the variables, for example, classified each study according to the level of training intervention. At the lowest level were short test-taking orientation sessions; at a somewhat higher level were longer coaching programs that included intensive, concentrated drill or "cramming" on sample test questions; and at the highest level was instruction in broad cognitive skills. In addition to test and program features, we coded methodological characteristics of the studies, features of the experimental populations, and publication features of the reports (Table 2).

Because different studies reported results on different scales, it was necessary to transform all the results to a common metric. The metric that we employed was the Effect Size or ES. This measure expresses differences between experimental and control scores in terms of standard-deviation units.

Results

The distribution of ES's was multi-modal in shape (Figure 1). One of the modes was at .1 standard deviations; another was at .4 standard deviations. The studies of coaching for the SAT were clustered tightly around the smaller mode; other studies were spread somewhat more loosely around the larger mode (Figure 2). Coaching programs for the SAT thus seemed to have different effects from coaching programs for other tests.

Further examination of the data showed that studies of SAT coaching were distinct from other studies in additional ways. Compared to other studies, the SAT studies were significantly more likely to involve long-term coaching, field-tested coaching programs, coaching by a commercial school, testing for a real-life educational decision, higher grade levels, pre/post research designs, and research carried out by ETS. Because SAT studies were different from

other studies both in features and in outcomes, we carried out all further analyses separately on SAT studies and on studies of other aptitude tests. Analysis of data from the total group would have produced misleading results.

Coaching for the SAT

All of the SAT studies employed both pretests and posttests. Improvement from initial to posttests averaged .36 standard deviations for the experimental groups and .21 standard deviations for the control groups. The effect of coaching alone, estimated from these 14 studies, was therefore equal to .36 minus .21, or .15 standard-deviation units.

None of the study features was significantly related to size of effect in the SAT studies. Effects were similar for SAT coaching programs of different durations and with different characteristics. Findings were also similar in groups of studies that used quite different methodologies or that employed distinctly different subject groups. And finally, findings were much the same for studies published in different ways and at different times.

Coaching for other Tests

Seventeen of the 24 studies of coaching for aptitude tests other than the SAT employed both pretests and posttests. Improvement from pretest to posttest averaged .76 standard deviations for the experimental groups and .25 standard deviations for the control groups. The effect of coaching alone, estimated from these 17 studies, was therefore equal to .76 minus .25, or .51 standard-deviation units. Studies that did not use pretests yielded a significantly lower estimate of the size of coaching effects. On the basis of all 24 studies, we estimated the average ES of coaching to be .43.

The use of a pretest in the experimental design turned out to be the only study feature significantly related to size of effect. Other study features were not significantly related to coaching outcomes.

Discussion

This meta-analysis showed that there are two distinct literatures on the effectiveness of coaching programs. The first is on the SAT and reports small effects from coaching. The second covers other aptitude tests and shows that coaching programs can have substantial effects.

The small SAT effects should not come as a surprise. Reviewers of the SAT coaching literature have repeatedly stated that the typical effect from SAT coaching is to raise

scores by between .1 and .2 standard-deviation units. Our results are consistent with these other findings, and yet we do not believe that the SAT is coach-proof. At least one well-designed study--by Evans and Pike (1973)--reported substantial coaching effects on the SAT. This study was carried out by ETS researchers who were thoroughly familiar with the SAT item pool and who developed special coaching materials for specific SAT item types. Other coachers have not been as familiar with SAT items because ETS security policies have until recently put SAT test forms out of their reach. Recent changes in ETS policies give the public much more access to SAT items and test forms, and it is possible that we will in the future see greater success for SAT coaching programs.

Some reviewers have speculated that program duration can explain much of the variation in the outcomes of studies of SAT effectiveness. Effective coaching programs for the SAT, they say, are long in duration while ineffective programs are short. Messick and Jungeblut (1981), in fact, have presented regression equations relating the logarithm of program length to the gain attributable to coaching. Using a pool of studies that differed slightly from Messick and Jungeblut's pool, we were unable to replicate their result. We do not believe therefore that the correlation that Messick and Jungeblut found between program duration and SAT effects is a robust one.

Our findings on coaching for other aptitude tests were similar to findings presented by Vernon in 1954. According to Vernon, practice and coaching can raise aptitude scores by about .6 standard deviations (or 8 to 9 points on an IQ scale). We found that the average combined effect of practice and coaching to be .76 standard deviations and the average effect attributable to coaching alone to be .4 standard deviations.

Studies that used a pretest yielded larger estimates of pure coaching effects than did other studies. In studies with a pretest, effects attributable to coaching averaged .51 standard deviations. In studies without pretests, effects of coaching averaged .27 standard deviations. It seems possible that the pretest acted to sensitize the students to the information presented in the coaching program. If so, a pretest may be an important component in any program designed to prepare students for aptitude tests.

We were not able to find other factors that influenced study results. Although this failure was disappointing, it was not unexpected. After examining results from numerous meta-analyses, Glass, McGaw, and Smith (1981) concluded reluctantly that the findings of contemporary research in the social sciences often fit together poorly, and that variation in study findings is only modestly predictable

from study characteristics. The results of our meta-analysis support this conclusion. Even with the use of objective tools for synthesis of findings, it was impossible to explain fully why coaching results differ as much as they do from study to study.

References

College Entrance Examination Board. (1968). Effects of coaching on Scholastic Aptitude Test scores. New York, NY: College Entrance Examination Board. (ERIC Document Reproduction Service, No. ED 169 130)

Evans, F. R., & Pike, L. W. (1973). The effects of instruction for three mathematics item formats. Journal of Educational Measurement, 10, 257-272.

Glass, G. V, McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills: Sage Publications.

Messick, S., & Jungeblut, A. (1981). Time and method in coaching for the SAT. Psychological Bulletin, 89, 191-216.

Vernon, P. E. (1954). Practice and coaching effects in intelligence tests. Educational Forum, 18, 269-280.

Note

The material in this paper is based upon work supported by the National Institute of Education under Grant No. G 81 0047. Any opinions, findings, and conclusions or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the National Institute of Education.

Table 1

Program and Test Features

Level of coaching

Duration of program in hours

Commercial vs. school program

Components of coaching program
--Testwiseness training
--Drill and practice
--Content teaching

Target test for coaching program
--Group vs. individual test
--Full test vs. subtest
--Teacher-made vs. standardized test
--SAT vs. other test

Table 2

Study Features

True vs. quasi-experiment

Pretest vs. posttest only design

Laboratory vs. field study

ETS-sponsored vs. other research

New vs. field-tested program

Grade level of students

Ability level of students

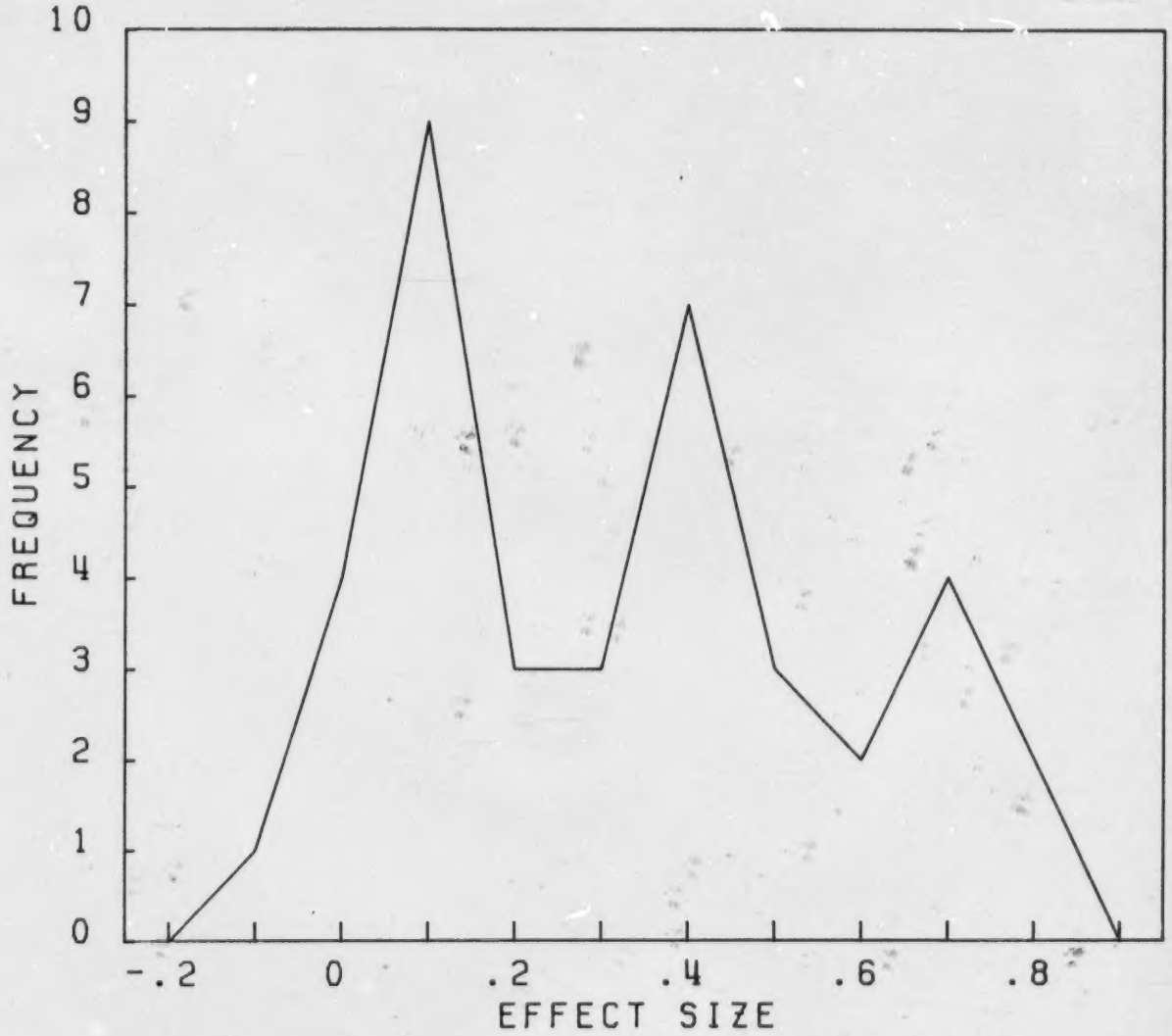
Source of report

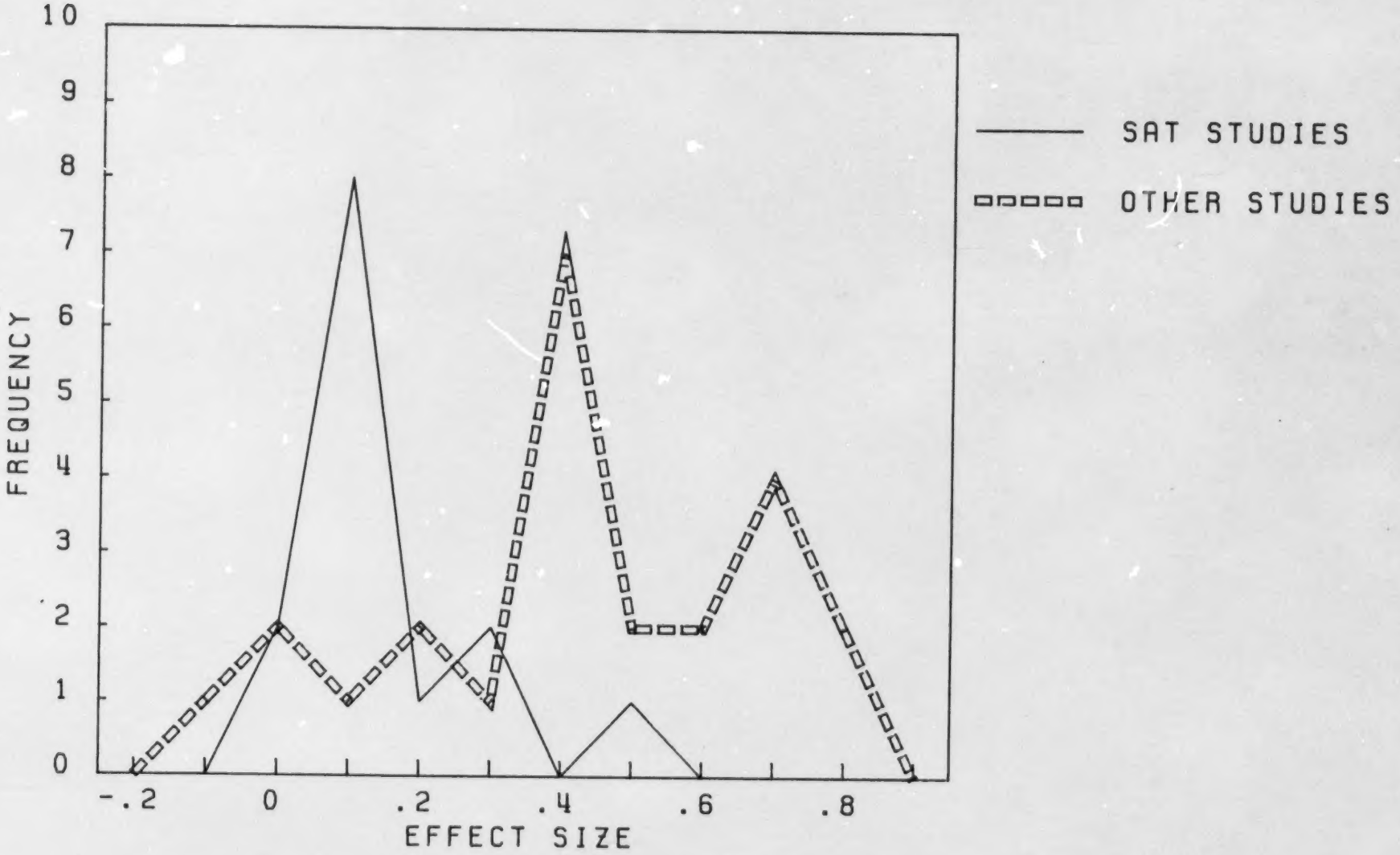
Year of report

Figure Captions

Figure 1. Distribution of coaching effects for 38 studies.

Figure 2. Distribution of coaching effects for 14 SAT studies and 24 studies of other aptitude tests.


10
9
8
7
6
5
4
3
2
1
0
FREQUENCY
-.2
0
.2
.4
.6
.8
EFFECT SIZE


10
9
8
7
6
5
4
3
2
1
0
FREQUENCY
-.2
0
.2
.4
.6
.8
EFFECT SIZE
SAT STUDIES
OTHER STUDIES